

This Provisional PDF corresponds to the article as it appeared upon acceptance. Fully formatted PDF and full text (HTML) versions will be made available soon.

## **The *Fasciola hepatica* genome: gene duplication and polymorphism reveals adaptation to the host environment and the capacity for rapid evolution**

*Genome Biology* (2015) 16:71

doi:10.1186/s13059-015-0632-2

Krystyna Cwiklinski (K.Cwiklinski@qub.ac.uk)  
John Pius Dalton (j.dalton@qub.ac.uk)  
Philippe J Dufresne (Philippe.Dufresne@inspq.qc.ca)  
James La Course (James.LaCourse@lstm.ac.uk)  
Diana JL Williams (williadj@liverpool.ac.uk)  
Jane Hodgkinson (jhodgkin@liverpool.ac.uk)  
Steve Paterson (s.paterson@liv.ac.uk)

Published online: 03 April 2015

**ISSN** 1465-6906

**Article type** Research

**Submission date** 10 November 2014

**Acceptance date** 13 March 2015

**Article URL** <http://dx.doi.org/10.1186/s13059-015-0632-2>

For information about publishing your research in BioMed Central journals, go to  
<http://www.biomedcentral.com/info/authors/>

# The *Fasciola hepatica* genome: gene duplication and polymorphism reveals adaptation to the host environment and the capacity for rapid evolution

Krystyna Cwiklinski<sup>1,2</sup>  
Email: K.Cwiklinski@qub.ac.uk

John Pius Dalton<sup>2,3</sup>  
Email: j.dalton@qub.ac.uk

Philippe J Dufresne<sup>3,4</sup>  
Email: Philippe.Dufresne@inspq.qc.ca

James La Course<sup>5</sup>  
Email: James.LaCourse@lstmed.ac.uk

Diana JL Williams<sup>1</sup>  
Email: williadj@liverpool.ac.uk

Jane Hodgkinson<sup>1</sup>  
Email: jhodgkin@liverpool.ac.uk

Steve Paterson<sup>6\*</sup>  
\* Corresponding author  
Email: s.paterson@liv.ac.uk

<sup>1</sup> Institute of Infection and Global Health, University of Liverpool, Liverpool, UK

<sup>2</sup> School of Biological Sciences, Medical Biology Centre, Queen's University of Belfast, Belfast, Northern Ireland, UK

<sup>3</sup> Institute of Parasitology, McGill University, Montreal, Quebec, Canada

<sup>4</sup> Institut National de Santé Publique du Québec, Montreal, Quebec, Canada

<sup>5</sup> Liverpool School of Tropical Medicine, Liverpool, UK

<sup>6</sup> Institute of Integrative Biology, University of Liverpool, Liverpool, UK

## Abstract

### Background

The liver fluke *Fasciola hepatica* is a major pathogen of livestock worldwide, causing huge economic losses to agriculture, as well as 2.4 million human infections annually.

## Results

Here we provide a draft genome for *F. hepatica*, which we find to be amongst the largest known pathogen genomes at 1.3Gb. This size cannot be explained by genome duplication or expansion of a single repeat element, and remains a paradox given the burden it may impose on egg production necessary to transmit infection. Despite the potential for inbreeding by facultative self-fertilization, substantial levels of polymorphism were found, which highlights the evolutionary potential for rapid adaptation to changes in host availability, climate change or to drug or vaccine interventions. Non-synonymous polymorphisms were elevated in genes shared with parasitic taxa, which may be particularly relevant for the ability of the parasite to adapt to a broad range of definitive mammalian and intermediate molluscan hosts. Large-scale transcriptional changes, particularly within expanded protease and tubulin families, were found as the parasite migrated from the gut, across the peritoneum and through the liver to mature in the bile ducts. We identify novel members of anti-oxidant and detoxification pathways and defined their differential expression through infection, which may explain the stage-specific efficacy of different anthelmintic drugs.

## Conclusions

The genome analysis described here provides new insights into the evolution of this important pathogen, its adaptation to the host environment and external selection pressures. This analysis also provides a platform for research into novel drugs and vaccines.

## Background

The digenean trematode *Fasciola hepatica* is one of the most important pathogens of domestic livestock and has a global distribution [1-4]. The disease, fasciolosis, results in huge losses to the agricultural industry associated with poor food conversion, lower weight gains, impaired fertility and reduced milk (cattle) and wool (sheep) production. Heavy, acute infections can result in death, particularly in sheep and goats. Economic losses attributable to *F. hepatica* infection have been estimated at more than US\$3 billion per annum worldwide [5,6], although even this estimate may be conservative as *F. hepatica* infection modulates its host's immune response and its ability to resist or eliminate common microbial pathogens [7,8]. Fasciolosis is also an important zoonosis in regions where agricultural management practices are less advanced, particularly in South America and North Africa [3,9]. It is estimated that between 2.4 and 17 million people are infected with this liver fluke worldwide, with a further 91 million people living at risk, resulting in fasciolosis being included on the WHO list of major neglected tropical diseases [1-3].

The zoonotic potential of *F. hepatica* is enabled by its remarkable ability to infect and mature in an extensive range of terrestrial mammals. Thus, while the typical definitive host for *F. hepatica* is one of many species of domestic or wild ruminant that ingest contaminated pasture (Figure 1), *F. hepatica* is also able to exploit disparate host species including humans and rodents, and has rapidly adapted to novel hosts such as llamas and kangaroos, which it has recently come into contact with in South America and Australia, respectively [10]. This is in contrast to most digenean trematodes, such as the human pathogen *Schistosoma mansoni*, which have a far more restricted host range. *F. hepatica* can also adapt rapidly to drug interventions and the emergence of resistance within *F. hepatica* populations to triclabendazole (TCBZ) is of major concern, since most drugs used against other digeneans

are only partly protective against *F. hepatica* [10]. TCBZ is also the only drug currently available that is able to protect livestock and humans against early stage juveniles, which cause significant pathology as they migrate through the liver. The ability of *F. hepatica* to adapt rapidly to novel hosts or to drug interventions is perhaps more remarkable given that *F. hepatica* is a hermaphrodite that can facultatively self-fertilize and so *F. hepatica* populations might be expected to lose genetic diversity through inbreeding that would be an essential basis for adaptation.

---

**Figure 1 *Fasciola hepatica* lifecycle.** (a) Graphical representation of the *F. hepatica* lifecycle (modified from [76]). *a1* Definitive host – host range includes cattle, sheep and humans. *a1.1* Parasite excysts in the intestine of the definitive host, releasing newly excysted juveniles (NEJ) that migrate across the intestinal wall, through the peritoneal cavity to the liver. *a1.2* NEJ migrate through the liver parenchyma, increasing in size to juvenile flukes as they migrate *a1.3* into the bile ducts *a1.4*, where they grow and develop into fully mature adults. *a2* Eggs are released in the faeces and develop on pasture. *a3* From each embryonated egg hatches a single miracidium, which infects the snail intermediate host (*Galba truncatula*). *a4* Within the snail the parasite undergoes a clonal expansion, developing through the sporocyst, rediae and cercariae lifecycle stages. *a5* Cercariae are released from the snail and encyst on vegetation as dormant metacercariae, which are ingested by the definitive host *a1*. (b-d) Graphical representation of the development of the parasite through the definitive host. (b) The parasite increases dramatically (~1000-fold) in size over the course of approximately 12 weeks, from NEJ to adult. (c) Expression of key enzymes of metabolism reveals how the growth of the parasite limits oxygen diffusion into the parasite tissue, switching from aerobic energy metabolism (Kreb's cycle; PK: pyruvate kinase; SD: succinate dehydrogenase) to aerobic acetate production (ME: malic enzyme) to anaerobic dismutation (PEPCK: phosphoenolpyruvate carboxykinase), as shown by the log fold change in expression between the lifecycle stages (expression is shown relative to metacercariae lifecycle stage). (d) In addition to the dramatic growth, maturation of the parasite occurs, with the fully mature adult digesting host blood, which provides the nutrient for massive egg production (approximately 20,000 eggs per day per parasite), as shown by the increased expression of the egg shell component, vitelline.

---

Here we provide a genome assembly for *F. hepatica* and assess genome-wide polymorphism and transcriptional profiles in order to identify key features of its genome that underlie its ability to migrate through different physiological environments, to parasitize different host species, and to respond rapidly to external selection pressures.

## Results and discussion

### A large genome with high gene polymorphism

A draft genome for *F. hepatica* was generated with an assembled length of approximately 1.3Gb (Table 1). The genome of *F. hepatica* is considerably larger than that of other sequenced digenean parasites – *Schistosoma spp* (363 - 397 Mb), *Clonorchis sinensis* (547 Mb) or *Opisthorchis viverrini* (634.5 Mb) [11-16] – and is one of the largest pathogen genomes sequenced to date. Genome size does not appear to be related to chromosome number among trematodes; *F. hepatica* has 10 pairs of chromosomes [17], *S. mansoni* and *O. viverrini* have 8 pairs and 6 pairs of chromosomes, respectively [18,19], but *C. sinensis*, also with a smaller genome than *F. hepatica*, has 28 pairs [20]. Comparative analysis with other

sequenced trematode species indicated that the mean number of exons per gene is comparable between species, but that mean exon and intron lengths tend to increase with genome size (Additional file 1: Table S1). Most core eukaryotic genes appeared as single copy evidenced by both CEGMA (Additional file 2: Table S2) and analysis of read coverage (Additional file 3: Figure S1), suggesting that the large genome size of *F. hepatica* has not arisen by genome duplication. At least 32% of the genome was estimated to consist of repetitive DNA, which is consistent with other trematode genomes [11-15]. The median repeat length was 26 bp (Additional file 4: Figure S2) and we observed retrotransposons, including 27Mbp of long terminal repeats and 59Mbp of long interspersed elements (LINEs); however, there was no obvious expansion of a single repeat element to account for the large genome size. A LINE RTE BovB repeat, previously found in ruminants, was observed distributed widely across the genome (at least 67,000 full or partial copies in total across approximately 30% of the scaffolds and totalling 28.1Mbp). This was not due to contamination by host DNA, since no other host sequence, such as sheep mitochondrial sequence, could be identified either in the assembly or in individual reads. BovB has previously been reported as exhibiting horizontal transfer between snakes and ruminants and its presence in *F. hepatica* suggests that transfer of BovB elements between disparate vertebrate taxa may be facilitated by digenian infection [21].

**Table 1 *Fasciola hepatica* assembly statistics**

Metric	Value
Scaffold N50	204Kbp (REAPR 155Kbp <sup>1</sup> )
Number of scaffolds $\geq 3$ Kbp	20,158
Number of scaffolds $\geq 1$ Kbp	45,354
Contig N50 ( $\geq 100$ bp)	9.7Kbp
Number of contigs ( $\geq 100$ bp)	254,014
Total assembly length	1.275Gbp
Total length of gaps	91.6Mbp
Repetitive content	32%
Number of RNAseq-supported gene models	22,676 (15,740 <sup>2</sup> )
Mean number of exons/gene	5.3
Mean exon size (95% range)	303 bp (36 bp - 1369 bp)
Mean intron size (95% range)	3.7Kbp (33 bp - 17.5Kbp)
Proportion CEGMA core eukaryotic genes found	90%

1 - N50 following breakage of some scaffolds at areas to low support. Both assemblies are available from ENA under project accession PRJEB6687.

2 - Number of non-overlapping, distinct genome intervals covered by RNAseq-supported gene models.

We investigated levels of polymorphism among *F. hepatica* genes by re-sequencing the genomes of individual fluke from each of five isolates, all from the UK. Substantial polymorphism among isolates was observed; 48% of genes exhibited at least one non-synonymous SNP and the level of non-synonymous nucleotide diversity,  $\pi$ , averaged across 21.8Mbp of coding sequence, was  $5.2 \times 10^{-4}$  (i.e. two randomly sampled sequences differed approximately every 1900bp). By comparison, this figure is higher than in humans [22], similar to most vertebrates [23] and, on limited data, smaller than some parasitic nematode populations [24]. Although *F. hepatica* is a self-fertilizing hermaphrodite, and so has the potential to inbreed and lose genetic diversity, our data show that *F. hepatica* populations, as a whole, harbour substantial genetic variation. A likely explanation is that parasite populations are typically large, often larger than that of their hosts, which greatly slows any enhanced effects of genetic drift caused by self-fertilization [25].

By analysing the distribution of genetic diversity amongst *F. hepatica* genes, we found higher non-synonymous polymorphism in genes shared with parasitic cestodes and digeneans relative to orthologs shared with the free-living turbellaria (Figure 2a and Additional file 5: Table S3 and Additional file 6: Table S4). These data suggest high adaptability in *F. hepatica* genes that mediate infection and survival in the host environment, which is consistent with *F. hepatica*'s ability to infect a range of both mammalian and molluscan hosts [3,9,10]. We then assessed whether high non-synonymous polymorphism was associated with particular biological functions and discovered a marked over-representation of biological processes associated with axonogenesis and chemotaxis among the top 1% quantile of polymorphic genes (Figure 2b and Additional file 7: Table S5 and Additional file 8: Table S6). These genes included cadherin, semaphorin, fascilin and rabconnectin, which are involved in cell adhesion and migration of neurons [26-28]. The high polymorphism observed in chemosensory and neural development pathways may relate to the challenge faced by *F. hepatica* in locating its snail host or in tissue migration in its vertebrate host, and with variation in host preference within parasite populations [29,30]. Such polymorphism may be particularly relevant for development of new anthelmintics targeting the parasite's neuromuscular system [31].

---

**Figure 2 Polymorphism within *Fasciola hepatica*.** (a) Levels of non-synonymous polymorphism for *F. hepatica* genes exhibiting orthology with *Clonorchis*, *Schistosoma*, *Schmidtea* or *Echinococcus* indicated within the phylogenetic tree. Numbers by branches refer to numbers of orthologous groups specifically shared by that branch; e.g. 464 orthologs are shared only between *Fasciola* and *Clonorchis*, a further 388 are also shared with *Schistosoma* but not with *Schmidtea* or *Echinococcus* etc. Branches not drawn to scale. Polymorphism is significantly ( $p < 0.001$ ) higher in *Fasciola* orthologs shared amongst the digenean and cestode parasites *Clonorchis*, *Schistosoma* and *Echinococcus* (green, red and blue crosses, respectively) relative to orthologs conserved with the turbellarian *Schmidtea* (black dots). (b-c) Directed acyclic graphs indicating over-representation of biological processes within the top 1% most polymorphic genes, based on non-synonymous diversity. Shaded boxes indicated significant ( $p < 0.01$ ) over-representation. Black and blue arrows indicate, respectively, 'is a' and 'part of' relationships between terms.

---

## Expression patterns from multi-gene families reveal important developmental host-parasite interactions

In order to understand how *F. hepatica* has adapted to survive within its vertebrate host, we characterised its developmental time-course of gene expression using RNAseq. Progressively more genes were differentially expressed, and with larger fold-changes, following initial infection and subsequent development in the host; i.e. associated with excysting of the metacercaria stage to form the newly excysted juvenile (NEJ), traversal of the intestinal wall and penetration of the liver as an immature tissue-migrating parasite (Figure 3a). Thereafter, the transition from these immature parasites to mature adults living in the bile duct was accompanied by further changes in the overall profile of differentially expressed genes, but not in the number of genes differentially expressed relative to metacercariae. Known aspects of *Fasciola* biology were recapitulated in our expression data. Thus, as the parasite grows, diffusion of oxygen into parasite tissue is limiting and our expression analysis confirmed the switch from aerobic to anaerobic metabolism (Figure 1b and c) [32]. Similarly, our data confirmed that the production of vitelline, the egg shell component, is limited to the mature parasites (Figure 1d).

---

**Figure 3 Expression of genes exhibiting at least a 16-fold difference in expression between any two developmental stages and grouped by hierarchical clustering. (a)** Heatmap with up-regulation in blue and down-regulation in red relative to metacercariae. Colors to left of heatmap correspond to different clusters. **(b)** Expression of genes within each cluster on a log<sub>2</sub> scale. The number of genes and the enrichment of biological processes are shown for the most specific terms within the gene ontology structure and with a significance of <0.1%. GO:0006508 Proteolysis is shown in brackets for clusters where it appears with a significance of 0.1% < p < 1%.

---

To explore novel processes associated with developmental changes, genes were clustered based on expression pattern (Figure 3b). Several biological processes were greatly down-regulated in maturing and adult parasites relative to metacercariae (Cluster 12, Figure 3), including molecules involved in cell adhesion (cadherins, integrins) and cytoskeletal proteins (talins) that could play an important role in sensing changes in the physiological environment and rapidly initiating excystment following ingestion by the definitive host. Clusters showing strongest patterns of differential expression (both up and down regulation) were markedly over-represented by peptidases and terms associated with regulation of peptidase activity (Clusters 2, 8, 10 and 11, Figure 3). Genes associated with *F. hepatica* structure, particularly protein polymerisation and microtubule based movement, such as tubulin, dynein and surface tegumental genes, were highly up-regulated in immature and adult fluke relative to earlier stages (Clusters 1 and 3, Figure 3b and Additional file 9: Table S7, Additional file 10: Table S8 and Additional file 11: Table S9). Our data show that, across the transcriptome as a whole, the strongest changes in expression were observed amongst different members of the multigenic protease and tubulin families, as detailed below.

Peptidases play essential roles in parasite infection, tissue migration and feeding [33]. *F. hepatica* is unique amongst helminth parasites because it relies almost exclusively on the secretion of papain-like cysteine peptidases (Clan A, family C1), classes cathepsins L (FhCL) and cathepsins B (FhCB) [33,34]. These two peptidase classes, however, have expanded and diverged to form multigenic families, all of which have been found within our assembly (Figure 4a and Additional file 12: Figure S3). Our data show that FhCL peptidases form clades with a total of 17 members and we found that each clade has a distinctive peptidolytic activity based on residues that make up the S2 sub-site of the active site (Additional file 12: Figure S3 [35,36]). Members within each clade exhibited a concerted pattern of expression that can be correlated with migration of the parasite through different tissues where they would encounter a different profile of macromolecular substrates. For example, FhCL3s are known to exhibit a unique collagenolytic activity required to disrupt the interstitial matrix and their secretion in high amounts enable NEJs to traverse the intestinal wall and penetrate the liver capsule [33,34,37]. Our data reveals that the FhCL3 clade has expanded to five members that are all abundantly expressed by NEJs and share critical residues (H61, W67 and V205) within the active site, suggesting that enlargement of this family has been driven by the need to produce plentiful collagenolytic peptidase at this critical time in the parasite's life cycle. Conversely, members of the FhCL1 clade, the largest with six members, were not expressed during the early migratory stages but all showed an increase in expression as the parasite matures (Figure 4b). Mature adults are obligate blood feeders [30,37] and secrete copious amounts of FhCL1 and, correspondingly, evolution of the FhCL1 active site (residues N61, L/F67 and L/M205) has resulted in loss of collagenolytic activity but a gain in the ability to effectively digest blood proteins, particularly hemoglobin, which are the mature parasite's prime source of nutrients required for egg production [30,37].

---

**Figure 4 Analysis of the cysteine proteases belonging to *F. hepatica* cathepsin L (FhCL) and cathepsin B (FhCB) clades and their activators, the asparaginyl endopeptidases/legumains.** (a) Representation of the number of genes identified for each FhCL clade and for the groups of FhCB and legumain genes, based on analysis by BLAST and manual annotation. The nucleotide identity is shown for each clade/ group of genes (Clustal Omega). (b-d) Graphical representation of the expression for FhCL, FhCB and legumain proteases across the *F. hepatica* lifecycle in reads per kilobase per million (RPKM). In the case of the FhCL gene expression, all the genes within the same clade showed a similar pattern of expression, and so are represented here as an average log RPKM. Likewise in the case where multiple gene models were identified for a particular gene, which showed a similar level of expression, are represented here as an average log RPKM.

---

Unlike the FhCL peptidases, the FhCB family consists of a single clade of seven members. Nevertheless, these peptidases were also temporally regulated (Figure 4c) such that three members (FhCB1, FhCB2 and FhCB3) exhibited parallel expression patterns to FhCL3 and were thus highly expressed in the NEJs and down-regulated as the parasites matured. These data suggest a concerted role for FhCLs and FhCBs in the early infection stage. Also, the constitution and specific expression of a family of peptidases, asparaginyl endopeptidases or legumains that are responsible for the processing of the inactive FhCL and FhCB zymogens to functional enzymes [38] suggest specific and important developmental roles for these peptidases (Figure 4d). Thus, legumain 1, which was the most highly expressed member in NEJs could be required for activation of the FhCL3 and FhCB 1/2/3 peptides at the time the parasite emerges from its encysted stage in the intestine and initiates infection. Furthermore, legumain 3, which was switched on late in parasite development, is the prospective candidate for activating the FhCL1, FhCL2 and FhCL5 in mature blood-feeding adult parasites.

Our data highlight tubulins as another multigenic family that exhibit amongst the highest log fold changes in expression throughout *F. hepatica* development. We identified the full complement of five  $\alpha$ -tubulin and six  $\beta$ -tubulin isotypes [39,40] and found duplication of  $\beta$ -tubulin isotype 3 (Figure 5a,b). Basal expression levels revealed two functional subsets of  $\beta$ -tubulin molecules; high constitutive expression of  $\beta$ -tubulin isotypes 2, 3a, 3b and 4 suggests they play an essential role in general microtubule structure and function, while a more specialized role is implicated for  $\beta$ -tubulin isotype 1 and 5, which were particularly expressed in the immature liver stages. Tubulins are known targets of benzimidazole (BZ) drugs [40,41], which include TCBZ, and the unique nature of the interaction of *F. hepatica* with TCBZ remains undefined. Protein similarity searches with the closely related species *C. sinensis* and *S. mansoni*, which are refractory to TCBZ, identified six and ten  $\beta$ -tubulin gene models, respectively. While *F. hepatica*  $\beta$ -tubulin isotypes 1–4 show >95% similarity with multiple sequences from these species, that observed for  $\beta$ -tubulin isotype 5 and 6 was <70% (Additional file 13: Table S10). Of the two BZ-derived drugs effective against *F. hepatica*, TCBZ targets immature and adult fluke whilst albendazole (ABZ), a widely used anti-nematode drug, kills only adult stages. Field isolates of *F. hepatica* that demonstrate resistance to ABZ remain susceptible to TCBZ suggesting independent modes of action [42] that may be related to the diverse  $\beta$ -tubulin isotypes found in this parasite.

---

**Figure 5 Expression of tubulin genes.** (a-b) Graphical representation of the expression of the tubulin genes across the *F. hepatica* lifecycle represented by  $\alpha$ -tubulins (a) and  $\beta$ -tubulins (b). The expression of each tubulin isotype (iso) is shown in reads per kilobase per million (RPKM), which were calculated as an average log RPKM value for the multiple transcriptome datasets at each lifecycle stage.

---



## Anti-oxidant and detoxification systems

We investigated the evolution of anti-oxidant systems in *F. hepatica*, which are essential for adaptation to the host environment. Thus, as the parasite rapidly develops and enters different aerobic/anaerobic environments, anti-oxidant systems are critical not only for the detoxification of reactive oxygen and nitrogen (ROS, RNS) generated by endogenous cellular metabolism but also as a frontline defence against superoxide and nitric oxide radicals released by host immune effector cells such as macrophages, eosinophils and neutrophils [43,44]. Parasitic platyhelminths express genes encoding superoxide dismutase (SOD) which dismutates superoxide radicals to  $H_2O_2$  but they do not possess catalase, the enzyme responsible for converting  $H_2O_2$  to water and oxygen (although a catalase gene is present in the genome of free-living flatworms, such as *Schmidtea mediterranea*). A catalase gene was not found in the *F. hepatica* genome, which is consistent with other parasitic platyhelminths, where the function of peroxide detoxification has been supplanted by the newly discovered thiol-dependent peroxiredoxin and its reducing partner thioredoxin [45]. The presence of the gene encoding the recently described thioredoxin glutathione reductase (TGR; Additional file 14: Table S11), together with the absence of distinct thioredoxin reductase and glutathione reductase genes verified that TGR is the sole reductive enzyme that links the thioredoxin-dependent and glutathione-dependent anti-oxidant defence systems in this parasite as observed in other platyhelminths [43,46,47]. The pivotal position of TGR between these two essential anti-oxidant systems makes it a promising target for the development of new anti-trematode drugs [46]. Included in the repertoire of *F. hepatica* anti-oxidants are genes encoding SOD, glutathione transferases (GSTs) and three fatty acid binding proteins (FABP; Additional file 14: Table S11). With the exception of the GSTs, we found that in *F. hepatica* each of these components is encoded by a single gene, which is in stark contrast to the expanded anti-oxidant gene families in the closely related trematodes, *O. viverrini* and *C. sinensis* [13,15]. Peroxiredoxin, GST and FABP are secreted by *F. hepatica* via non-classical secretory pathways, perhaps as cargoes of exosomes [37], into the host circulation where they influence host immune responses by recruiting and activating M2 macrophages [7,48,49] and suppressing dendritic cell activity [50]. These immunomodulatory effects contribute to the establishment of an immune suppressive environment, which aids parasite survival and the development of chronic disease.

Defence against chemical toxins, including those generated by the immune response, is essential in allowing helminths to adapt to, and survive in, the host environment. This is mediated in large part by the three-phase drug detoxification pathway; Phase I (activation), Phase II (conjugation) and Phase III (efflux). This pathway also acts to reduce drug activity and/or bioavailability and has been linked with TCBZ resistance [41,51-53]. Until now only indirect evidence existed for the presence of Phase I cytochrome P450 (CYP 450) genes in *F. hepatica* [54]. Here, we report that the *F. hepatica* genome contains two Phase I CYP 450 gene models, one monooxygenase and one reductase; both are present in *S. mansoni* but only the reductase is found in *C. sinensis*. Basal expression levels did not reveal significant changes throughout development (Figure 6). Our analysis reveals that of the Phase II GSTs, the cytosolic GSTs were represented in the *F. hepatica* genome by at least seven separate genes comprising four different classes; four mu, one omega, one sigma and a putative novel zeta class. Zeta class GSTs demonstrate different activities to other GSTs and their role in helminths is poorly understood [55]. Comparison with *S. mansoni* and *C. sinensis* revealed this zeta class GST is unique to *F. hepatica*. Within the Phase III pathway [56,57], we identified 18 ATP-Binding Cassette (ABC) and five Multidrug and Toxic compound Extrusion (MATE) transporters in *F. hepatica*, similar to that found in *S. mansoni* and *C.*

*sinensis*. Although no significant developmental changes in gene expression were observed for the cytosolic GSTs, we found that two microsomal GSTs and several putative multidrug resistance proteins were up-regulated in the immature and adult stages (Figure 6 and Additional file 15: Table S12) supporting the importance of broad-spectrum detoxification Phases in mediating defence against host-immune-mediated toxic assault and targeting via anti-parasitic drugs.

---

**Figure 6 Expression of detoxification genes.** Expression of Phase I, II and II detoxification genes through development relative to metacercariae.

---

## Conclusions

The *F. hepatica* genome is one of the largest pathogen genomes sequenced to date but we found no evidence of genome duplication or repeat expansion to explain this. Why this large genome should have evolved is unclear, especially given that its replication may be energetically costly or slow cell division [58]. For a parasite, such as *Fasciola*, that relies on the production of large numbers of eggs to facilitate transmission, one would expect strong selection against the accumulation of junk DNA if a large genome imposed a cost on egg production. It is possible that much of the non-coding portion of the *F. hepatica* genome is involved in gene regulation [59], and the size of *F. hepatica*'s genome may be related to its complex life-cycle and variety of developmental morphs. If so, however, it would be difficult to explain why the *F. hepatica* genome is around three times the size of the *Schistosoma* genome [11], which has a similar life-cycle. The large genome of *F. hepatica* therefore remains a paradox for which we may have to wait for comparative genome sequencing, currently underway [60], across other platyhelminth taxa to provide an answer.

The ability of *F. hepatica* to infect and survive in different tissue environments as it migrates from the intestine, through the liver and into the bile ducts is underpinned by gene duplication. Thus, our results show that, across the whole transcriptome, the strongest patterns of differential expression were observed among members of protease and tubulin gene families, and, in the case of proteases, these can be associated with changes in the active site and substrate specificity. While gene duplication appears to be a key process of adaptation to the parasitic life-style used by many helminths, it is notable that different helminth taxa have arrived at different evolutionary solutions. Comparison between *F. hepatica* and the bile-dwelling liver flukes *C. sinensis* and *O. viverrini*, shows the expansion of the anti-oxidant SOD families [13] and cathepsin F protease families [13,61] in *C. sinensis* and *O. viverrini* but not in *F. hepatica*, and conversely the expansion of the cathepsin L family in the latter species only, which suggests differences in how these related parasites tackle life within the same environment. Differences in the specificity or developmental expression of detoxification genes, such as members of the ABC and MATE families, or of the tubulin gene family, may be important in understanding why *F. hepatica* responds so markedly different to drugs at different stages of development compared to other digeneans. The exclusive activity of TCBZ against *F. hepatica*, and the potency of praziquantel to all these other digeneans except *F. hepatica*, points to a uniqueness in this parasite that needs to be resolved [37].

*Fasciola hepatica* is a highly adaptable parasite, evidenced by its ability to infect novel hosts, and it is notable that our results reveal high levels of polymorphism in genes specific to parasitic digeneans. Diversity within *F. hepatica* populations at genes important for the host-

parasite interface may underpin a high evolutionary potential for *F. hepatica* to respond to changes in host availability or to other selection pressures. Similarly, the broad geographic range of *F. hepatica* would suggest that it is able to adapt to different climatic conditions and, in this respect, *F. hepatica* may also be able to respond to exploit changes in climate in temperate regions, such as the UK, where warmer and wetter winters favour transmission and increased prevalence [62]. *F. hepatica* is seen to rapidly develop drug resistance to TCBZ [52] and the standing genetic diversity that we find across the genome suggests that it harbours the potential to evolve resistance to any novel drug treatment, which compounds the difficulty of controlling *F. hepatica* given the shortage of drugs effective against the juvenile stages. Nevertheless, the availability of a genome for *F. hepatica* that we provide, plus the characterisation of early molecular events in infection should help support the development of novel drugs and vaccines, particularly against the migrating juveniles that cause much of the pathology. For example, RNAi-mediated knockdown of cathepsin L and B expression has been shown to prevent NEJs from crossing the intestine [45] and vaccine trials with cathepsin L1 have provided partial protection against infection and pathology [63]. The exploitation of a broader range of targets within the *F. hepatica* genome is now a priority given the widespread prevalence of resistance to TCBZ within *F. hepatica* populations. The commercial importance for agriculture to develop a replacement treatment to TCBZ may stimulate new treatments that could be translated to other important digenean parasites, including those of humans.

## Methods

### Source of parasite material

Adult parasites from each of five isolates were used for genome sequencing: 1) FhepLivSP, from the laboratory maintained Shrewsbury isolate (Ridgeway Research, UK); 2) FhepLivS1, a clonal line derived from the Shrewsbury population; 3) and 4) FhepLivR1 and R3, clonal lines derived from two isolates recovered from sheep in Northern England naturally infected with *F. hepatica* and 5) FhepLivR2, a clonal line derived from a *F. hepatica* population from naturally infected sheep in South West Wales. For RNA sequencing, the following were used: 1) metacercariae and newly excysted juveniles (NEJ) at 1, 3 and 24 hours post excystment from a North American isolate (Baldwin Aquatics Inc., USA). 21 day old juvenile flukes were recovered from mice infected with the same isolate; 2) an adult parasite recovered from the bile ducts of cattle naturally infected with *F. hepatica* in Uruguay. All animal work was conducted with ethical approval from the Universities of Liverpool (UK) and McGill (Canada).

### Sequencing

Approximately 10 µg of DNA from a single, adult fluke taken from the FhepLivS1 strain was used to prepare Illumina TruSeq fragment libraries and 26Gbp of 2 x 250 bp reads generated on an Illumina MiSeq (mean insert sizes 470 bp and 580 bp). Further individuals of FhepLivS1 were used to prepare Nextera Mate-Pair libraries (3Kbp and 10Kbp) and approximately 60 m 2x 100 bp Illumina reads from each library were generated. Approximately 200 µg of DNA prepared from several fluke of the FhepLivSP isolate, was used to construct 2Kbp, 5Kbp and 8Kbp mate-pair libraries, which were sequenced either on an Illumina GAII or HiSeq 2000. For each of the parasite isolates FhepLivSP, FhepLivS1, FhepLivR1, FhepLivR2 and FhepLivR3, DNA was isolated from a single adult; a fragment

library was prepared and sequenced on a single lane of an Illumina HiSeq 2000 to yield approximately 24Gbp of sequence (Centre for Genomic Research, Liverpool, UK). For RNA sequencing, Illumina TruSeq RNA libraries were prepared from biological replicates of metacercariae (3 replicates), NEJ 1 hour (2 replicates), NEJ 3 hours (2 replicates), NEJ 24 hours (2 replicates), Juveniles 21 days (1 replicate) and Adult (1 replicate) (Genome Quebec, Montreal, Canada).

## **Assembly and annotation**

Illumina MiSeq reads were trimmed to  $Q \geq 30$  and adaptors removed using Sickel and Perl and assembled using Newbler (Roche GS-Assembler v2.6) with flags set for large genome and a heterozygote sample. Mate-pair reads were first mapped to these contigs using Bowtie2 [64] to remove duplicates and wrongly orientated reads, and scaffolded into contigs using SSPACE [65]. Gap filling was achieved using GapFiller for 2x 250 bp and 2x 100 bp paired-end reads and run for three iterations (available as ENA accessions LN627018-LN647175). RNAseq data were mapped to scaffolds within the assembled genome greater than 3Kbp using TopHat2 to identify transcribed regions and splice junctions. These, together with RNAseq data assembled using Trinity and *S. mansoni* protein sequence, were passed to the MAKER pipeline [66] to predict genes. Repeatmasker, Windowmasker and Dustmasker were used to identify repetitive regions. CEGMA v2.4 [67], which searches for 248 highly conserved genes, was used to assess the completeness of genome with the settings for vertebrates to allow long introns. REAPR [68] was used to assess the quality of scaffolding within the assembly and to produce an alternative, more conservative assembly by splitting scaffolds at locations with lower support (available as ENA accessions LN736597-LN774150). Homologs of *F. hepatica* predicted protein sequences were identified within UniProt using BLAST and functional domains identified using InterPro. InParanoid and MultiParanoid [69] were used to identify ortholog clusters from *Schistosoma mansoni* (v3.1.16), *Clonorchis sinensis* (v3.7), *Schmidtea mediterranea* and *Echinococcus multilocularis* (v29042013) predicted proteins [11-15,70].

## **Gene expression analysis**

RNAseq libraries were mapped to MAKER gene models using TopHat2 [71] and read counts extracted using htseq-count. Genes with a sum of at least five reads across all libraries were analysed for differential expression in edgeR [72] using a negative binomial model of successive developmental stages relative to metacercariae and with tagwise dispersion estimated from all samples. Hierarchical clustering, based on model coefficients, was used to group differentially expressed genes by similarity of expression into 12 clusters and hypergeometric tests used to test for over-representation of gene ontology terms within each cluster relative to the whole gene-set.

## **Genetic diversity analysis**

Reads for each isolate were mapped to the genome using Bowtie2 and resulting bam files passed together to the GATK pipeline [73] for local realignment and SNP calling. SNP calls within genes were filtered by score  $>100$ , FS  $<60$  and combined coverage between 10 and 250. Because gene duplicates collapsed within the assembly might give erroneously high diversity for some genes, genes were excluded with a median coverage greater than 213 and by heuristic scoring of SNPs appearing as heterozygotes in all samples. IGV was used to manually assess the success of filtering parameters. Levels of nucleotide diversity for

different classes of SNPs were calculated for all genes. To identify the most polymorphic genes, a generalised linear model with a Poisson distribution was fitted to the number of non-synonymous SNPs as the response variable versus number of non-synonymous sites as a quadratic function. This accounts for the distribution of SNP counts within a gene, the fact that longer genes have the potential to have more SNPs and the possibility that genes of different lengths may evolve at different rates. Genes were classified according to their conservation across platyhelminths by the presence of orthologs retained across increasing taxonomic scales. This was preferred to assessing conservation on the basis of non-synonymous versus synonymous substitution ratios between species, since synonymous sites were saturated by multiple substitutions at such broad taxonomic scales. The robustness of the generalised linear model was tested by randomly sampling equal proportions of genes from each quartile of the length distribution for each class of conserved gene, to ensure that no bias was introduced if the length of a gene was correlated to its level of conservation. From the generalised linear model, the most polymorphic genes were identified having residuals in the top 1% quantile and hypergeometric tests were used to test for over-representation of GO terms within these highly polymorphic genes relative to the gene-set as a whole.

## **Discovery and characterisation of gene families**

Discovery of *F. hepatica* gene families was carried out using BLAST analysis (NCBI v2.2.27 & v2.2.29), with available published gene sequences of interest (Additional file 16: Table S13), followed by manual annotation. Comparative analysis was carried out using the closely related trematode genome sequence datasets: *Clonorchis sinensis* and *Schistosoma mansoni*. Sequence alignment and phylogenetic analysis was carried out using Clustal Omega [74] and MEGA5 [75] respectively.

## **Data availability**

Data are freely available from WormBase ParaSite and the European Nucleotide Archive under accessions LN627018-LN647175 (assembly data), PRJEB6687 (genomic read data) and PRJEB6904 (transcriptomic read data).

## **Competing interests**

The authors declare that they have no competing interests.

## **Author contributions**

JH, JPD, SP and DW conceived the study, obtained funding and contributed resources. KC, JH, JPD and SP designed the study, interpreted the data and wrote the manuscript, with substantial input from PD, DW and JLC. SP performed genome assembly, automated annotation and analysis of polymorphism and differential expression. KC and JLC annotated and analysed gene families. KC, JH, PD and DW generated and prepared parasite material for sequencing. All authors read and approved the final manuscript.

## Acknowledgements

We are grateful to Katherine Allen for producing clonal lines of liver fluke and Catherine Hartley, University of Liverpool and Paula Martin and Oliver Gladstone, Ridgeway Research, UK for technical assistance in producing parasite material. We are grateful for sequencing support provided by Margaret Hughes and Suzanne Kay at the Centre for Genomic Research, University of Liverpool, and Mathieu Bourgey, Genome Quebec, Canada. This study was funded by grants from the Biotechnology and Biological Sciences Research Council, UK (BB1002480/1) to JH, SP, DW and JLC, and a Ministère Économie, Innovation et Exportation (MEIE), Québec, award and a European Research Council Advanced Grant (HELIVAC, 322725) to JPD.

## References

1. Keiser J, Utzinger J. Food-borne trematodiasis. Clin Microbiol Rev. 2009;22:466–83.
2. Fürst T, Keiser J, Utzinger J. Global burden of human food-borne trematodiasis: a systematic review and meta-analysis. Lancet Infect Dis. 2012;12:210–21.
3. Robinson MW, Dalton JP. Zoonotic helminth infections with particular emphasis on fasciolosis and other trematodiasis. Philos Trans R Soc Lond B Biol Sci. 2009;364:2763–76.
4. Videnova K, Mackay DKJ. Availability of vaccines against major animal diseases in the European Union. Rev Sci Tech. 2012;31:971–8.
5. Spithill TW, Carmona C, Piedrafita D, Smooker PM. Prospects for immunoprophylaxis against *Fasciola hepatica* (liver fluke). In: Targets, Screens, Drugs and Vaccines. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2012. p. 465–84.
6. Piedrafita D, Spithill TW, Smith RE, Raadsma HW. Improving animal and human health through understanding liver fluke immunology. Parasite Immunol. 2010;32:572–81.
7. Dalton JP, Robinson MW, Mulcahy G, O'Neill SM, Donnelly S. Immunomodulatory molecules of *Fasciola hepatica*: Candidates for both vaccine and immunotherapeutic development. Vet Parasitol. 2013;195:272–85.
8. Claridge J, Diggle P, McCann CM, Mulcahy G, Flynn R, McNair J, et al. *Fasciola hepatica* is associated with the failure to detect bovine tuberculosis in dairy cattle. Nat Commun. 2012;3:853.
9. García HH, Moro PL, Schantz PM. Zoonotic helminth infections of humans: echinococcosis, cysticercosis and fascioliasis. Curr Opin Infect Dis. 2007;20:489–94.
10. Mas-Coma S, Bargues M-D, Valero MA. Fascioliasis and other plant-borne trematode zoonoses. Int J Parasitol. 2005;35:1255–78.
11. Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, Cerqueira GC, et al. The genome of the blood fluke *Schistosoma mansoni*. Nature. 2009;460:352–65.

12. Protasio AV, Tsai IJ, Babbage A, Nichol S, Hunt M, Aslett MA, et al. A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. PLoS Negl Trop Dis. 2012;6:e1455.
13. Young ND, Nagarajan N, Lin SJ, Korhonen PK, Jex AR, Hall RS, et al. The *Opisthorchis viverrini* genome provides insights into life in the bile duct. Nat Commun. 2014;5:e4378.
14. Zhou Y, Zheng H, Chen Y, Zhang L, Wang K, Guo J, et al. The *Schistosoma japonicum* genome reveals features of host-parasite interplay. Nature. 2009;460:345–51.
15. Huang Y, Chen W, Wang X, Liu H, Chen Y, Guo L, et al. The carcinogenic liver fluke, *Clonorchis sinensis*: new assembly, reannotation and analysis of the genome and characterization of tissue transcriptomes. PLoS One. 2013;8:e54732.
16. Young ND, Jex AR, Li B, Liu S, Yang L, Xiong Z, et al. Whole-genome sequence of *Schistosoma haematobium*. Nat Genet. 2012;44:221–5.
17. Sanderson AR. Maturation and probable gynogenesis in the liver fluke, *Fasciola hepatica* L. Nature. 1953;172:110–2.
18. Hirai H, Hirai Y, LoVerde PT. Evolution of sex chromosomes ZW of *Schistosoma mansoni* inferred from chromosome paint and BAC mapping analyses. Parasitol Int. 2012;61:684–9.
19. Komalamisra C. Chromosomes and C-banding of *Opisthorchis viverrini*. Southeast Asian J Trop Med Public Health. 1999;30:576–9.
20. Park G-M, Im K-I, Huh S, Yong T-S. Chromosomes of the liver fluke, *Clonorchis sinensis*. Korean J Parasitol. 2000;38:201.
21. Adelson DL, Raison JM, Edgar RC. Characterization and distribution of retrotransposons and simple sequence repeats in the bovine genome. Proc Natl Acad Sci. 2009;106:12855–60.
22. Li W-H, Sadler LA. Low nucleotide diversity in man. Genetics. 1991;129:513–23.
23. Gayral P, Melo-Ferreira J, Glémin S, Bierne N, Carneiro M, Nabholz B, et al. Reference-free population genomics from next-generation transcriptome data and the vertebrate–invertebrate gap. PLoS Genet. 2013;9:e1003457.
24. Skuce P, Stenhouse L, Jackson F, Hypša V, Gilleard J. Benzimidazole resistance allele haplotype diversity in United Kingdom isolates of *Teladorsagia circumcincta* supports a hypothesis of multiple origins of resistance by recurrent mutation. Int J Parasitol. 2010;40:1247–55.
25. Anderson TJC, Romero-Abel ME, Jaenike J. Genetic structure and epidemiology of *Ascaris* populations: patterns of host affiliation in Guatemala. Parasitol. 1993;107:319–34.
26. Scheiffele P. Cell-cell signaling during synapse formation in the CNS. Annu Rev Neurosci. 2003;26:485–508.

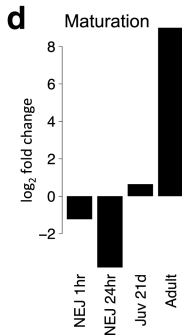
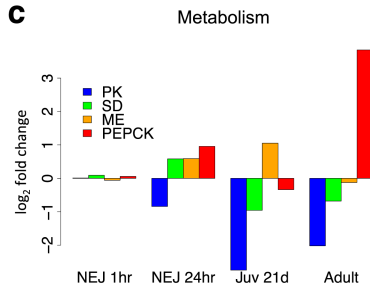
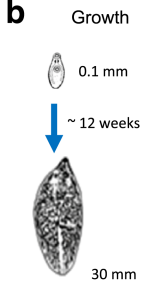
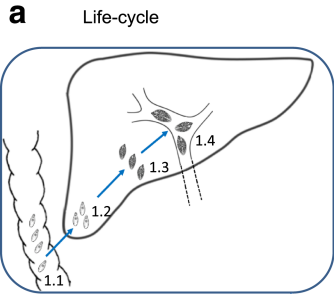
27. Zito K, Fetter RD, Goodman CS, Isacoff EY. Synaptic clustering of Fascilin II and Shaker: essential targeting sequences and role of Dlg. *Neuron*. 1997;19:1007–16.
28. Tuttle AM, Hoffman TL, Schilling TF. Rabconnectin-3a regulates vesicle endocytosis and canonical wnt signaling in zebrafish neural crest migration. *Plos Biol*. 2014;12:e1001852.
29. Kalbe M, Haberl B, Haas W. Snail host finding by *Fasciola hepatica* and *Trichobilharzia ocellata*: compound analysis of “miracidia-attracting glycoproteins.”. *Exp Parasitol*. 2000;96:231–42.
30. Andrews SJ. The life cycle of *Fasciola hepatica*. In: Dalton JP, editor. *Fasciolosis*. Wallingford, Oxfordshire, UK: CABI; 1999.
31. McVeigh P, Atkinson L, Marks NJ, Mousley A, Dalzell JJ, Sluder A, et al. Parasite neuropeptide biology: Seeding rational drug target selection? *Int J Parasitol Drugs Drug Resist*. 2012;2:76–91.
32. Tielens AGM. Metabolism. In: Dalton JP, editor. *Fasciolosis*. Wallingford, Oxfordshire, UK: CABI; 1999
33. McVeigh P, Maule AG, Dalton JP, Robinson MW. *Fasciola hepatica* virulence-associated cysteine peptidases: a systems biology perspective. *Microbes Infect*. 2012;14:301–10.
34. Robinson MW, Tort JF, Lowther J, Donnelly SM, Wong E, Xu W, et al. Proteomics and phylogenetic analysis of the cathepsin L protease family of the helminth pathogen *Fasciola hepatica*: expansion of a repertoire of virulence-associated factors. *Mol Cell Proteomics*. 2008;7:1111–23.
35. Turk V, Stoka V, Vasiljeva O, Renko M, Sun T, Turk B, et al. Cysteine cathepsins: from structure, function and regulation to new frontiers. *Biochim Biophys Acta*. 2012;1824:68–88.
36. Corvo I, O’Donoghue AJ, Pastro L, Pi-Denis N, Eroy-Reveles A, Roche L, et al. Dissecting the active site of the collagenolytic cathepsin L3 protease of the invasive stage of *Fasciola hepatica*. *PLoS Negl Trop Dis*. 2013;7:e2269.
37. Robinson MW, Menon R, Donnelly SM, Dalton JP, Ranganathan S. An integrated transcriptomics and proteomics analysis of the secretome of the helminth pathogen *Fasciola hepatica*: proteins associated with invasion and infection of the mammalian host. *Mol Cell Proteomics*. 2009;8:1891–907.
38. Dalton JP, Brindley PJ, Donnelly S, Robinson MW. The enigmatic asparaginyl endopeptidase of helminth parasites. *Trends Parasitol*. 2009;25:59–61.
39. Ryan LA, Hoey E, Trudgett A, Fairweather I, Fuchs M, Robinson MW, et al. *Fasciola hepatica* expresses multiple  $\alpha$ - and  $\beta$ -tubulin isoforms. *Mol Biochem Parasitol*. 2008;159:73–8.

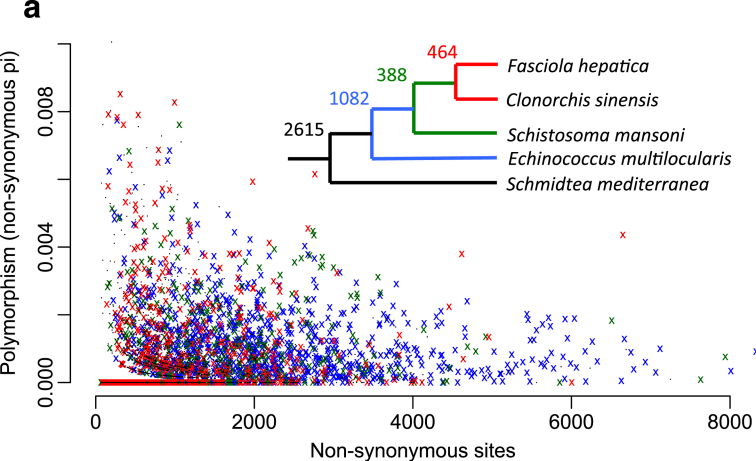


40. Fuchs MA, Ryan LA, Chambers EL, Moore CM, Fairweather I, Trudgett A, et al. Differential expression of liver fluke  $\beta$ -tubulin isotypes at selected life cycle stages. *Int J Parasitol*. 2013;43:1133–9.
41. Devine C, Brennan GP, Lanusse CE, Alvarez LI, Trudgett A, Hoey E, et al. Inhibition of triclabendazole metabolism in vitro by ketoconazole increases disruption to the tegument of a triclabendazole-resistant isolate of *Fasciola hepatica*. *Parasitol Res*. 2011;109:981–95.
42. Sanabria R, Ceballos L, Moreno L, Romero J, Lanusse C, Alvarez L. Identification of a field isolate of *Fasciola hepatica* resistant to albendazole and susceptible to triclabendazole. *Vet Parasitol*. 2013;193:105–10.
43. Williams DL, Bonilla M, Gladyshev VN, Salinas G. Thioredoxin glutathione reductase-dependent redox networks in platyhelminth parasites. *Antioxid Redox Signal*. 2013;19:735–45.
44. Lu J, Holmgren A. The thioredoxin antioxidant system. *Free Radic Biol Med*. 2014;66:75–87.
45. McGonigle L, Mousley A, Marks NJ, Brennan GP, Dalton JP, Spithill TW, et al. The silencing of cysteine proteases in *Fasciola hepatica* newly excysted juveniles using RNA interference reduces gut penetration. *Int J Parasitol*. 2008;38:149–55.
46. Prast-Nielsen S, Huang H-H, Williams DL. Thioredoxin glutathione reductase: Its role in redox biology and potential as a target for drugs against neglected diseases. *Biochim Biophys Acta*. 2011;1810:1262–71.
47. Maggioli G, Silveira F, Martín-Alonso JM, Salinas G, Carmona C, Parra F. A recombinant thioredoxin-glutathione reductase from *Fasciola hepatica* induces a protective response in rabbits. *Exp Parasitol*. 2011;129:323–30.
48. Robinson MW, Hutchinson WF, Dalton JP, Donnelly S. Peroxiredoxin: a central player in immune modulation. *Parasite Immunol*. 2010;32:305–13.
49. Donnelly S, Stack CM, O'Neill SM, Sayed AA, Williams DL, Dalton JP. Helminth 2-Cys peroxiredoxin drives Th2 responses through a mechanism involving alternatively activated macrophages. *FASEB J*. 2008;22:4022–32.
50. Dowling DJ, Hamilton CM, Donnelly S, La Course J, Brophy PM, Dalton J, et al. Major secretory antigens of the helminth *Fasciola hepatica* activate a suppressive dendritic cell phenotype that attenuates Th17 cells but fails to activate Th2 immune responses. *Infect Immun*. 2010;78:793–801.
51. Wilkinson R, Law CJ, Hoey EM, Fairweather I, Brennan GP, Trudgett A. An amino acid substitution in *Fasciola hepatica* P-glycoprotein from triclabendazole-resistant and triclabendazole-susceptible populations. *Mol Biochem Parasitol*. 2012;186:69–72.
52. Brennan GP, Fairweather I, Trudgett A, Hoey E, McCoy, McConville M, et al. Understanding triclabendazole resistance. *Exp Mol Pathol*. 2007;82:104–9.

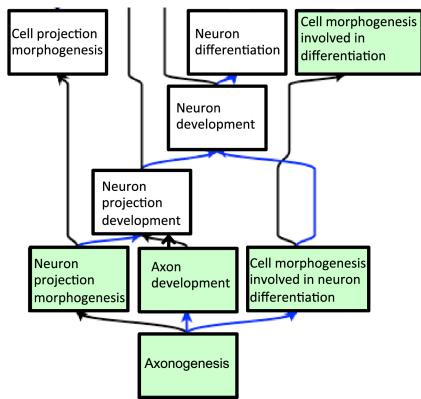
53. Virkel G, Lifschitz A, Sallovitz J, Ballent M, Scarcella S, Lanusse C. Inhibition of cytochrome P450 activity enhances the systemic availability of triclabendazole metabolites in sheep. *J Vet Pharmacol Ther.* 2009;32:79–86.
54. Devine C, Brennan GP, Lanusse CE, Alvarez LI, Trudgett A, Hoey E, et al. Potentiation of triclabendazole action in vivo against a triclabendazole-resistant isolate of *Fasciola hepatica* following its co-administration with the metabolic inhibitor, ketoconazole. *Vet Parasitol.* 2012;184:37–47.
55. Morphey RM, Eccleston N, Wilkinson TJ, McGarry J, Perally S, Prescott M, et al. Proteomics and *in silico* approaches to extend understanding of the glutathione transferase superfamily of the tropical liver fluke *Fasciola gigantica*. *J Proteome Res.* 2012;11:5876–89.
56. Reed MB, Panaccio M, Strugnell RA, Spithill TW. Developmental expression of a *Fasciola hepatica* sequence homologous to ABC transporters. *Int J Parasitol.* 1998;28:1375–81.
57. Young ND, Hall RS, Jex AR, Cantacessi C, Gasser RB. Elucidating the transcriptome of *Fasciola hepatica* -- A key to fundamental and biotechnological discoveries for a neglected parasite. *Biotechnol Adv.* 2010;28:222–31.
58. Pagel M, Johnstone RA. Variation across species in the size of the nuclear genome supports the junk-dna explanation for the c-value paradox. *Proc Biol Sci.* 1992;249:119–24.
59. Consortium TEP. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489:57–74.
60. Holroyd N, Sanchez-Flores A. Producing parasitic helminth reference and draft genomes at the Wellcome Trust Sanger Institute. *Parasite Immunol.* 2012;34:100–7.
61. Kang J-M, Bahk Y-Y, Cho P-Y, Hong S-J, Kim T-S, Sohn W-M, et al. A family of cathepsin F cysteine proteases of *Clonorchis sinensis* is the major secreted proteins that are expressed in the intestine of the parasite. *Mol Biochem Parasitol.* 2010;170:7–16.
62. Fox NJ, White PCL, McClean CJ, Marion G, Evans A, Hutchings MR. Predicting impacts of climate change on *Fasciola hepatica* risk. *PLoS One.* 2011;6:e16126.
63. Zafra R, Pérez-Écija RA, Buffoni L, Moreno P, Bautista MJ, Martínez-Moreno A, et al. Early and late peritoneal and hepatic changes in goats immunized with recombinant cathepsin 11 and infected with *Fasciola hepatica*. *J Comp Pathol.* 2013;148:373–84.
64. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357–9.
65. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics.* 2011;27:578–9.
66. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, et al. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 2007;18:188–96.

67. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*. 2007;23:1061–7.
68. Hunt M, Kikuchi T, Sanders M, Newbold C, Berriman M, Otto TD. REAPR: a universal tool for genome assembly evaluation. *Genome Biol*. 2013;14:R47.
69. Alexeyenko A, Tamas I, Liu G, Sonnhammer ELL. Automatic clustering of orthologs and inparalogs shared by multiple proteomes. *Bioinformatics*. 2006;22:e9–15.
70. Blythe MJ, Kao D, Malla S, Rowsell J, Wilson R, Evans D, et al. A dual platform approach to transcript discovery for the planarian *Schmidtea mediterranea* to establish RNAseq for stem cell and regeneration biology. *PLoS ONE*. 2010;5:e15617.
71. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14:R36.
72. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. 2012;40:4288–97.
73. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491–8.
74. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7:539–9.
75. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28:2731–9.
76. Centers for Disease Control and Prevention 'Laboratory identification of parasitic diseases of public health concern'. <http://www.cdc.gov/dpdx/fascioliasis/index.html>.

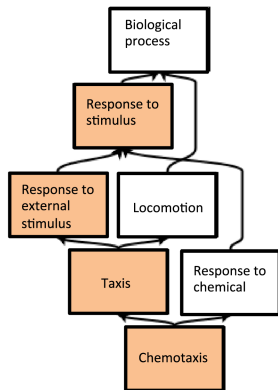


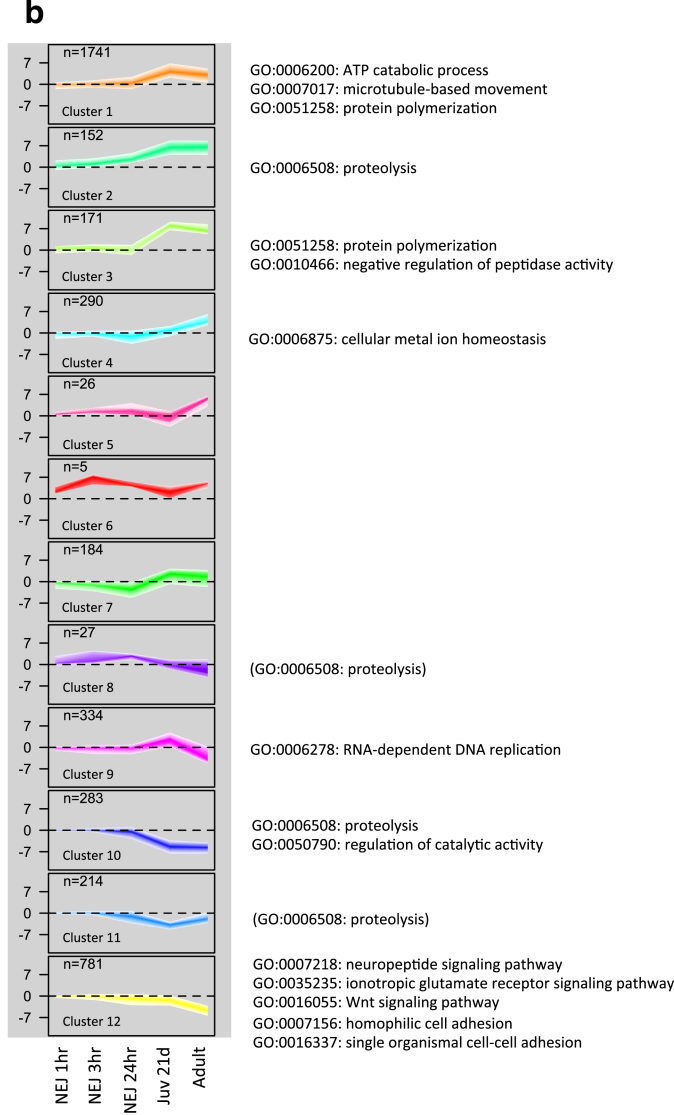
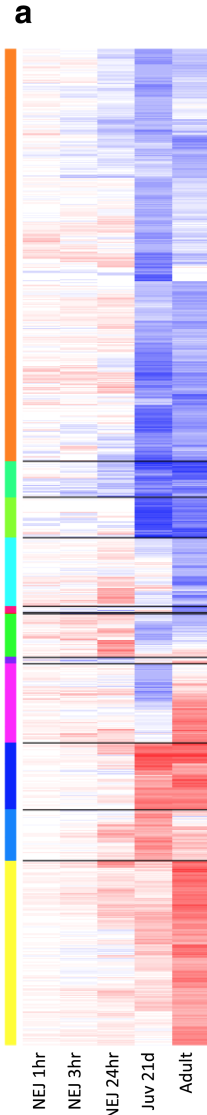


**b**



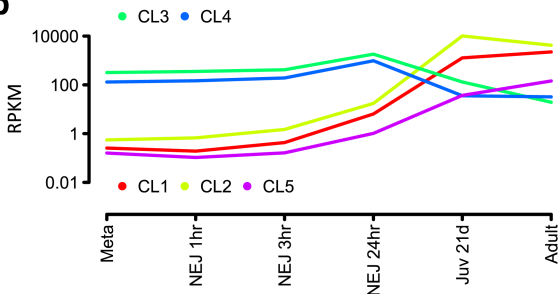
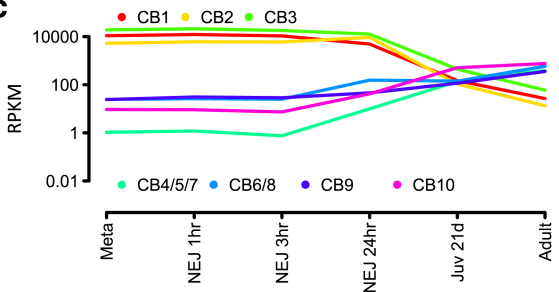
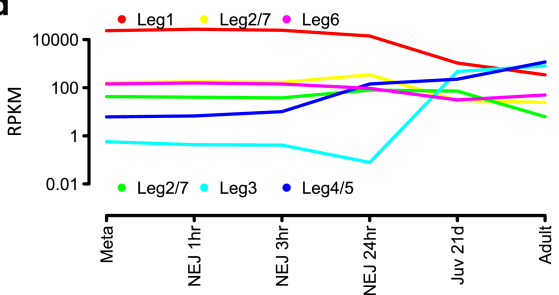
**c**

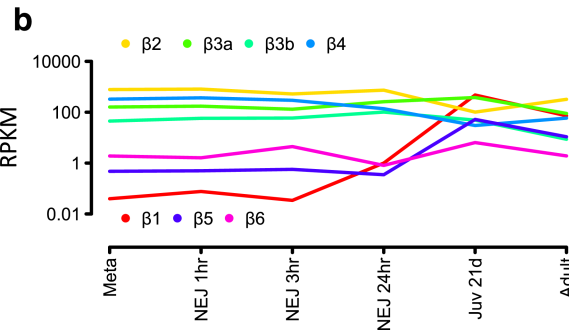
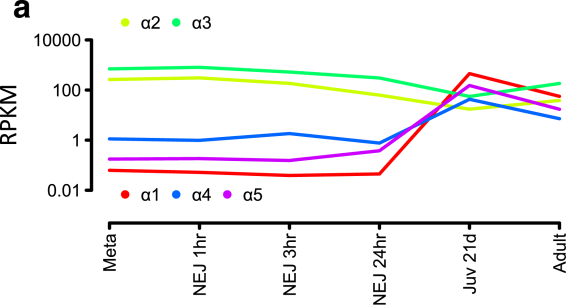




**a**

	Cathepsin L					Cathepsin B	Legumains
	Clade 1	Clade 2	Clade 3	Clade 4	Clade 5		
No. of genes	6	1	5	2	3	7	6
Nucleotide identity (%)	87-97	/	82-98	89	78-83	57-73*	53-77^

**b****c****d**





$\log_2$  fold change



-5

0

5

Phase I

Phase II

Phase III

NEJ 1hr

NEJ 3hr

NEJ 24hr

Juv 21d

Adult

## Additional files provided with this submission:

**Additional file 1: Table S1.** Comparison of the *F. hepatica* genome with other parasitic trematode genomes (42kb)

<http://genomebiology.com/content/supplementary/s13059-015-0632-2-s1.xlsx>

**Additional file 2: Table S2.** Summary statistics of CEGMA analysis (32kb)

<http://genomebiology.com/content/supplementary/s13059-015-0632-2-s2.xlsx>

**Additional file 3: Figure S1.** Distribution of average read coverage across gene models (104kb)

<http://genomebiology.com/content/supplementary/s13059-015-0632-2-s3.pdf>

**Additional file 4: Figure S2.** Distribution of repeat length within the *Fasciola hepatica* genome (44kb)

<http://genomebiology.com/content/supplementary/s13059-015-0632-2-s4.pdf>

**Additional file 5: Table S3.** Number of *Fasciola hepatica* ortholog clusters identified in *Clonorchis sinensis*, *Schistosoma mansoni*, *Schmidtea mediterranea* and *Echinococcus multilocularis* (9kb)

<http://genomebiology.com/content/supplementary/s13059-015-0632-2-s5.xlsx>

**Additional file 6: Table S4.** Association between number of non-synonymous SNPs per gene and orthology with other platyhelminths. Showing the output of a generalised linear model with a quasi-poisson error structure (51kb)

<http://genomebiology.com/content/supplementary/s13059-015-0632-2-s6.xlsx>

**Additional file 7: Table S5.** Over-representation of gene ontology (GO) terms in the top 1% quantile polymorphic genes, as estimated from number of non-synonymous SNPs controlled for length (9kb)

<http://genomebiology.com/content/supplementary/s13059-015-0632-2-s7.xlsx>

**Additional file 8: Table S6.** Top 1% most polymorphic genes (94kb)

<http://genomebiology.com/content/supplementary/s13059-015-0632-2-s8.xlsx>

**Additional file 9: Table S7.** Over-representation of GO Terms associated with each co-expression cluster (11kb)

<http://genomebiology.com/content/supplementary/s13059-015-0632-2-s9.xlsx>

**Additional file 10: Table S8.** Identity of gene models within each of 12 co-expression clusters, including annotation (4646kb)

<http://genomebiology.com/content/supplementary/s13059-015-0632-2-s10.txt>

**Additional file 11: Table S9.** Differential expression of genes presented in Figure 3. Values are log2 fold changes relative to metacercaria (1158kb)

<http://genomebiology.com/content/supplementary/s13059-015-0632-2-s11.txt>

**Additional file 12: Figure S3.** Analysis of the clades representing the cathepsin L cysteine proteases (167kb)

<http://genomebiology.com/content/supplementary/s13059-015-0632-2-s12.pdf>

**Additional file 13: Table S10.** Comparative analysis of the *F. hepatica*  $\beta$  tubulin genes with *C. sinensis* and *S. mansoni*, based on protein similarity (12kb)

<http://genomebiology.com/content/supplementary/s13059-015-0632-2-s13.xlsx>

**Additional file 14: Table S11.** *F. hepatica* antioxidant analysis (37kb)

<http://genomebiology.com/content/supplementary/s13059-015-0632-2-s14.xlsx>

**Additional file 15: Table S12.** Differential expression of detoxification genes (57kb)

<http://genomebiology.com/content/supplementary/s13059-015-0632-2-s15.xlsx>

**Additional file 16: Table S13.** Table of *Fasciola* sequences used for BLAST analysis (13kb)

<http://genomebiology.com/content/supplementary/s13059-015-0632-2-s16.xlsx>